

4 Methoden

4.1 Patienten, Ethik, Validität und Testauswahl

4.1.1 Einschlusskriterien für Patienten

Alle geeigneten teilnahme willigen Patienten mit Carotisendarterektomie sollten konsekutiv in diese Studie aufgenommen werden. Von der Studie ausgeschlossen wurden Patienten, bei denen vorab oder im Verlauf der psychologischen Tests erkennbar war, dass sie nicht in der Lage waren, die Tests durchzuführen. Jede Testsitzung begann mit einem Mini-Mental-Status-Demenztest, um Patienten mit Demenzverdacht (< 24 Punkte im Test) und Patienten mit mangelnden Fertigkeiten der deutschen Sprache und Schrift zu erkennen und von der weiteren Testdurchführung auszuschließen. Außerdem sollte so der Demenzstatus aller so ausgeschlossenen und teilnehmenden Patienten dokumentiert werden, um damit einer Forderung von Rao (2001) nachzukommen.

4.1.2 Ethik und Datenschutz

Die Patienten erhielten eine schriftliche Patienteninformation (siehe Seite 106) und wurden zusätzlich mündlich über die Ziele und Methoden der Studie und über den Datenschutz informiert. Die teilnehmenden Patienten unterschrieben eine Einverständniserklärung (siehe Seite 107).

Die ärztliche Leitung und regelmäßige wissenschaftliche Betreuung der Studie übernahm der neurologische Oberarzt Herr PD Dr. Andreas Hartmann, der auch die neurologischen und Ultraschall-Untersuchungen durchführte.

Die Erhebung und Speicherung der Daten wurde gemäß dem Berliner Datenschutzgesetz dem Datenschutzbeauftragten der Klinik gemeldet.

Personenbezogene Daten (Name, Geburtsdatum, Adresse, Telefonnummer) und Daten zur Untersuchungsteilnahme (Notizen, Angaben wegen Terminen und Gründen für Nichtteilnahme von Patienten) wurden in einer nur mir zugänglichen kennwortgesicherten Access-Datenbank gespeichert. Über die Patientenummer konnten diese Daten bei Bedarf den Messergebnissen zugeordnet werden.

Zur Dokumentation der Testdurchführung wurden die durch eine Patientenummer gekennzeichneten Antwortbögen und die Einverständniserklärungen der Patienten (mit Patientenummer, Name und Adresse) in getrennten Aktenordnern aufbewahrt.

Die Richtlinien der Ethikkommission der Klinik bezüglich Information der Patienten, Freiwilligkeit der Studienteilnahme und Datenschutz wurden eingehalten. Die Ethikkommission wurde über das Studiendesign und die Bedingungen der Studiendurchführung informiert und stellte in der Sitzung vom 30. Januar 2001 die Unbedenklichkeit fest.

4.1.3 Validität der Veränderungsmessungen

Die Validität der Studie hätte durch systematische Störeinflüsse gefährdet werden können. Vier wesentliche Probleme waren hier zu bedenken:

- Erholungseffekte bei Patienten, die kurz vor der Operation einen Schlaganfall erlitten hatten (Lunn 1999, p.77).
- Übungseffekte bei Testwiederholungen (Lunn 1999, p.79).
- Vorübergehende wahrscheinlich negative Folgen der Operation (Irvine 1998, p.200).
- Auswirkungen der unterschiedlichen Lebensumstände der Patienten vor und nach der Operation.

Effekte einer fortgesetzten Erholung von einem früheren Schlaganfall auf die Testleistung könnten nur durch einen ausreichend langen Abstand der Operation zu dem ischämischen Ereignis ausgeschlossen werden. Aus ethischen Gründen war eine Verzögerung der Operation aber nicht zulässig. Zur Kontrolle von möglichen Erholungseffekten sollte deshalb ein Vergleich von Hirninfarkt- und TIA-Patienten durchgeführt werden.

Übungseffekte wurden kontrolliert durch Auswahl von möglichst übungsresistenten Tests und durch den Einsatz von Tests mit unterschiedlichen Parallelförmigkeiten für jede Messung wie von Irvine (1998) empfohlen. Außerdem sollte die Follow-up Untersuchung mindestens drei Monate nach der Operation liegen. Dadurch würden auch mögliche vorübergehende Folgen der Operation ausgeschlossen: Alle vier Studien, die kognitive Verschlechterungen nach CEA berichteten, hatten das Follow-up innerhalb des ersten Monats nach der Operation durchgeführt (siehe Tabelle 1 auf Seite 25).

Am schwierigsten kontrollierbar erscheinen mir mögliche Beeinträchtigungen der kognitiven Leistungen der Patienten zum Zeitpunkt der Erstuntersuchung vor CEA durch die Angst vor der Operation und die Umstände des Klinikaufenthaltes. Zum Zeitpunkt der Nachuntersuchung erschienen die Patienten dann ambulant und ohne diese Belastungen. Dieses Problem wird übrigens in keinem der drei Reviews erwähnt (Irvine 1998; Lunn 1999; Rao 2001). Eine Kontrolle der Auswirkungen emotionaler Faktoren auf die kognitive Leistung sollte durch Messung der Depressivität und des emotionalen Status und Korrelation mit den kognitiven Leistungen erfolgen.

4.1.4 Kontrollgruppe

Eine Randomisierung der Patienten nach dem ersten Test auf eine Behandlungsgruppe mit CEA und eine Kontrollgruppe ohne CEA wäre ideal, um die im letzten Abschnitt besprochenen möglichen Probleme zu kontrollieren. Diese Möglichkeit ist aber ethisch nicht vertretbar, da man nach dem aktuellen Kenntnisstand den Patienten mit hochgradigen symptomatischen Stenosen die operative Behandlung empfehlen muss (Irvine 1998, p.197; Kelly 1980, p.743).

Alternativ dazu könnte eine Kontrollgruppe von Patienten herangezogen werden, die einer anderen ähnlich schwerwiegenden Operation unterzogen werden und die nach Alter, Ge-

schlecht, Bildungsniveau und Testergebnissen zu den CEA-Patienten passen. In vergleichbaren Studien wurde häufig auf Kontrollgruppen verzichtet und wo Kontrollgruppen verwendet wurden, haben sie sich in Aufwand, Vergleichbarkeit und Interpretation als problematisch erwiesen (Lunn 1999; Rao 2001). Zur Vermeidung des mindestens doppelten Aufwandes bei möglicherweise nur geringem Nutzen, habe ich mich deshalb für meine Studie gegen eine Kontrollgruppe entschieden.

Ein bisher nur selten verfolgter, aber alternativer Ansatz, der den Einsatz einer Nicht-CEA-Kontrollgruppe erübrigt, ist der Vergleich von nach CEA hämodynamisch verbesserten und unveränderten Patienten bezüglich Veränderungen der kognitiven Leistung (siehe Seite 27).

Bei etwa einer Hälfte der Patienten können vor CEA hämodynamische Veränderungen im Gehirn nachgewiesen werden (Eckstein 2003; Hartmann 2000). Nach CEA normalisiert sich die Hämodynamik dieser Patienten meistens (Kim 2002).

Wie im Kapitel Fragestellung (siehe Seite 30ff) dargestellt, ist nach Normalisierung der Hämodynamik eine stellenweise Verbesserung der Gehirn-Perfusion zu erwarten. Die verbesserte Perfusion könnte zu verbesserten kognitiven Leistungen führen.

In den Reviews von Irvine (1998) und Rao (2001) wird die Einbeziehung von cerebralen Blutflussmessungen in weiteren Studien zu kognitiven Veränderungen nach CEA gefordert und Lunn (1999, p.78) verlangt, Subgruppen von CEA-Patienten miteinander zu vergleichen.

Meine Arbeit soll diesen Anregungen und Überlegungen folgen, indem die kognitiven Veränderungen von intrakraniell hämodynamisch verbesserten und unveränderten Patienten nach CEA verglichen werden.

4.1.5 Testauswahl

Irvine (1998, p.201) fordert den Einsatz von möglichst anwenderfreundlichen kognitiven Leistungstests, weil die meist älteren CEA-Patienten geringer belastbar sind. Murkin (1995) hat Probleme und Qualitätsmerkmale von Tests neuropsychologischer Veränderungen nach Herzoperationen zusammengestellt. Diese Forderungen wurden in der vorliegenden Studie bei der Testauswahl berücksichtigt: Es wurde auf einfache Durchführbarkeit, hohe Objektivität, mittlere Schwierigkeit, Vermeidung von Deckeneffekten, geringe Übungseffekte, hohe Teststärke, hohes Skalenniveau mit differenzierter Skala, und geringe emotionale Belastung der Patienten geachtet.

Alle ausgewählten Testverfahren sollten weitgehend selbsterklärend sein, um eine möglichst objektive Durchführung zu gewährleisten. Zur besseren Vergleichbarkeit sollten möglichst in der Forschung etablierte und evaluierte Tests verwendet werden.

Im Bereich der für den Alltag relevanten Fähigkeiten sollten kognitive Fähigkeiten gemessen werden, die für die Patienten bei der Bewältigung ihres Alltagslebens eine große Bedeutung haben. Die Ausführungsgeschwindigkeit sollte nur eine geringe Rolle spielen, da es wichtiger ist, bestimmte Aufgaben überhaupt selbstständig bewältigen zu können als diese besonders schnell, effektiv oder in großer Zahl zu erledigen.

Neben einem Test der allgemeinen Intelligenz sollte das Gedächtnis in mehreren Gegenstandsbereichen getestet werden:

1. Lebens- und Langzeitgedächtnis anhand eines Wortschatztests,
2. Kurzzeitgedächtnis und Konzentration anhand eines Zahlengedächtnistests,
3. Gedächtnis für Gesichter.

Die emotionale Lage sollte durch einen Depressivitätstest und durch einen Test zur Selbstbeurteilung der Stimmung, Leistungsfähigkeit und Gesundheit ermittelt werden. Damit beim Selbstbeurteilungstest die Fragen auch einzeln ausgewertet werden können, sollten diese ein gut differenziertes Ordinalskalen-, besser Intervallskalenniveau erreichen.

Die Suche nach geeigneten Tests umfasste vor allem das Sichten der Literatur vergleichbarer Studien (siehe Tabelle 1 auf Seite 25), Suchmaschinenrecherchen im Internet und die Befragung einer praktizierenden Testpsychologin, eines Psychiaters und des Leiters der psychologischen Testbibliothek der Freien Universität Berlin (siehe Kapitel 8.5, Seite 133).

Alle so gefundenen Gesichtergedächtnis-Tests testeten gleichzeitig das Namensgedächtnis und waren in der Durchführung zu aufwändig. Deshalb wurde ein Gesichtergedächtnis-Test (GGT, siehe Seiten 37 und 112) selbst konstruiert.

Außerdem wurde ein eigener Test zur Selbstbeurteilung der Stimmung, Leistungsfähigkeit und Gesundheit mit visuellen Analogskalen (VAS, siehe Seite 45) konstruiert, weil für diesen Bereich kein Test gefunden wurde, der auch bei den einzelnen Fragen ein hohes Skalenniveau erreichte.

Der Hauptnachteil solcher Eigenkonstruktionen ist, dass es keine Referenzwerte gibt. Da es in dieser Studie jedoch um Leistungsänderungen geht, wird jeder Patient im Follow-up an seinem eigenen Test vor der Operation gemessen. Ein zweiter Nachteil von Eigenkonstruktionen ist, dass Schwierigkeit, Reliabilität und Validität nicht vorab bekannt sind. Durch wohlüberlegte Zusammenstellung sollte jedoch eine ausreichende Reliabilität und Validität gewährleistet werden können. Die Schwierigkeit des Gesichter-Gedächtnis-Tests wurde im Verlauf einiger Vorversuche mit Familienangehörigen, Studenten und Kollegen und einem Patienten noch etwas erhöht, um beobachtete Deckeneffekte zu vermeiden.

Um bei den Leistungstests Retesteffekte zu minimieren, wurden, wie von Irvine (1998, p.199) empfohlen, Verfahren ausgewählt, für die Paralleltests vorliegen (Mehrfachwahl-Wortschatz-Test und Gesichter-Gedächtnis-Test) oder es wurde eine nach Aufgabenschwierigkeit parallelisierte Testteilung vorgenommen (Standard Progressive Matrices). Um systematische Wirkungen von Unterschieden beider Testformen auszuschließen, wurde die Reihenfolge der Anwendung der Paralleltests für jeden Patienten randomisiert.

Außerdem wurden die routinemäßig durchgeführten neurologischen Untersuchungen dokumentiert und ausgewertet.

4.2 Erhobene Daten

Im Folgenden werden die verwendeten Testverfahren zitiert, beschrieben und abgebildet oder es wird auf den Anhang verwiesen. Eine vollständige Wiedergabe aller verwendeten Tests würde sehr viel Platz beanspruchen. Der SPM wurde zum Beispiel in zwei Versionen zu je 32 Blatt verwendet. Der Gesichter-Gedächtnis-Test in zwei Versionen zu je 18 Blatt. Bei mehrseitigen Tests wird deshalb auf eine vollständige Wiedergabe verzichtet. Stattdessen wird jeweils nur die Referenz und ein Ausschnitt wiedergegeben, so dass es möglich ist, das jeweilige Verfahren nachzuvollziehen.

4.2.1 Kognitive Leistung

Die kognitive Leistung wurde vor der Operation und im Follow-up mit folgenden Tests gemessen:

Mini-Mental-Status-Test (MMST) Demenztest

Nach der Zusammenstellung von Bickel (2000) steigt die Prävalenz von Demenzerkrankungen in Deutschland von 1,2% bei den 65 bis 69jährigen bis auf 34,6% bei den über 90jährigen. Etwa ein sechstel der Schlaganfall-Patienten hatte schon vor ihrem Schlaganfall eine Demenz (Henon 1997). Nach Schlaganfällen ist das Demenz-Risiko um ein vielfaches erhöht (Erkinjuntti 2004) und etwa 25% der Patienten nach Schlaganfall entwickeln eine Demenz und noch mehr zeigen kognitive Einschränkungen ohne Demenz. Auch für die übrigen Patienten nach Schlaganfall ist das Demenzrisiko erhöht (Rundek 2004). Trotzdem fehlen in manchen Studien zu kognitiven Einschränkungen von Carotis-Patienten Angaben zu Demenz-Diagnosen wie Rao (2001) kritisiert.

Durch den Einsatz des Mini-Mental-Status-Tests (siehe Bild 4 auf Seite 37) wurde der häufig geäußerten Forderung nach Kontrolle des Demenzstatus bei Studien zur kognitiven Leistung von CEA-Patienten gefolgt (Irvine 1998, p.201).

Der Mini-Mental-Status-Test ist ein Instrument zur Erfassung kognitiver Störungen bei älteren Personen. Der Test besteht aus 30 alltagsnahen Fragen und Aufgaben zu Orientierung, Aufnahmefähigkeit, Aufmerksamkeit, Rechnen, Gedächtnis, Sprache, Ausführung einer Anweisung, Lesen, Schreiben und Zeichnen nach einer geometrischen Figur (Folstein 1975). Erreichbar sind 0 bis 30 Punkte. Normale Probanden erreichen im Durchschnitt 27,6 Punkte. Bei weniger als 24 Punkten besteht die Möglichkeit einer Demenz und es wird eine genauere Diagnostik empfohlen. Die Interrater-Reliabilität beträgt $r=0,827$, die Retestreliabilität in 24 Stunden Abstand $r=0,887$ (Testzentrale Göttingen 2004).

Mini-Mental Status Test (MMST)			
Name	Datum	Score	Score
A. Orientierung			
Zeit (z.B. Welchen Tag haben wir heute?)	1. Jahr	1	
	2. Jahreszeit	1	
	3. Datum	1	
	4. Wochentag	1	
	5. Monat	1	
Ort (z.B. Wo sind wir?)	6. Land/Staat	1	
	7. Bundesland	1	
	8. Stadt/Ortschaft	1	
	9. Klinik/Praxis/Altersheim	1	
	10. Stockwerk	1	
	Summe (max. 10):	—	
B. Merkfähigkeit			
Der Untersucher nennt folgende drei Gegenstände und fordert den Patienten auf, die Begriffe zu wiederholen (1 Punkt für jede richtige Antwort). Der Untersucher wiederholt die Wörter solange, bis der Patient alle drei gelernt hat (höchstens 6 Wiederholungen).	1. >Auto<	1	
	2. >Blume<	1	
	3. >Kerze<	1	
	Summe (max. 3):	—	
C. Aufmerksamkeit und Rechenfähigkeit			
Von 100 an sind jeweils 7 abzuziehen. Falls ein Rechenfehler gemacht wird, so wird nur ein Fehler gegeben.	1. >93<	1	
	2. >86<	1	
	3. >79<	1	
	4. >72<	1	
	5. >65<	1	
ODER Falls der Patient die Aufgabe nicht durchführen kann oder will, „Radio“ rückwärts buchstabieren lassen: O-I-D-A-R	1. O	1	
	2. I	1	
	3. D	1	
	4. A	1	
	5. R	1	
	Summe (max. 5):	—	
D. Erinnerungsfähigkeit			
Der Untersucher fragt nach den drei zuvor genannten Wörtern.	1. >Auto<	1	
	2. >Blume<	1	
	3. >Kerze<	1	
	Summe (max. 3):	—	
E. Sprache			
Der Untersucher zeigt zwei Gegenstände und fordert den Patienten auf, sie zu benennen.	1. Armbanduhr	1	
	2. Bleistift	1	
Der Untersucher fordert den Patienten auf, nachzusprechen.	3. „Sie leih ihm kein Geld mehr“	1	
Der Untersucher lässt den Patienten folgendes Kommando befolgen.	4. „Nehmen Sie dieses Blatt in die rechte Hand“	1	
	5. „Falten Sie es in der Mitte“	1	
	6. „Legen Sie es auf den Boden“	1	
Der Untersucher bittet den Patienten,	7. Die Anweisung auf der Rückseite zu befolgen	1	
Der Untersucher dreht das Blatt um und fordert den Patienten auf,	8. einen vollständigen Satz zu schreiben (Rückseite).	1	
Der Untersucher lässt den Patienten die auf der Rückseite vorgegebene Figur malen (1 Punkt, wenn alle Seiten und Winkel stimmen und die sich überschneidenden Linien ein Viereck bilden).	9. Nachzeichnen (Rückseite)	1	
	Summe (max. 9):	—	
	Gesamtsumme bitte oben eintragen	—	

Bitte schließen Sie die Augen!

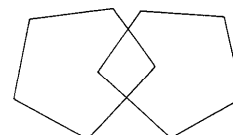


Bild 4: Mini-Mental-Status-Test: Links die Aufgaben, die der Versuchsleiter mündlich stellt, rechts das für die letzten drei Aufgaben dem Patienten vorzulegende Blatt (hier verkleinert).

Gesichter-Gedächtnis-Test (GGT)

Die Fähigkeit, Gesichter wiederzuerkennen ist für das menschliche Zusammenleben wichtig. Der in dieser Studie eingesetzte Gesichter-Gedächtnis-Test (GGT, siehe Bild 5) zum Wiedererkennen von Gesichtern wurde selbst entwickelt. Zur Vermeidung von Übungseffekten beim Follow-up nach der Operation wurden zwei Testversionen X und Y konstruiert. Welche Testversion ein Patient vor und welche er nach der Operation erhielt, wurde randomisiert.

Das Testmaterial einer Testversion bestand aus 9 Kartons (1A, 1B, 1C, 2A, 2B, 2C, 3A, 3B, 3C) für drei Messzeitpunkte (1 bis 3) innerhalb der Testsitzung: 1 = sofort, 2 = nach etwa 20 Minuten, 3 = nach etwa 40 Minuten je nach Dauer der dazwischen liegenden Tests. Vor jeder Testsitzung wurde randomisiert, welche Testkartenummer zu welchem der drei Messzeitpunkte verwendet wurde.

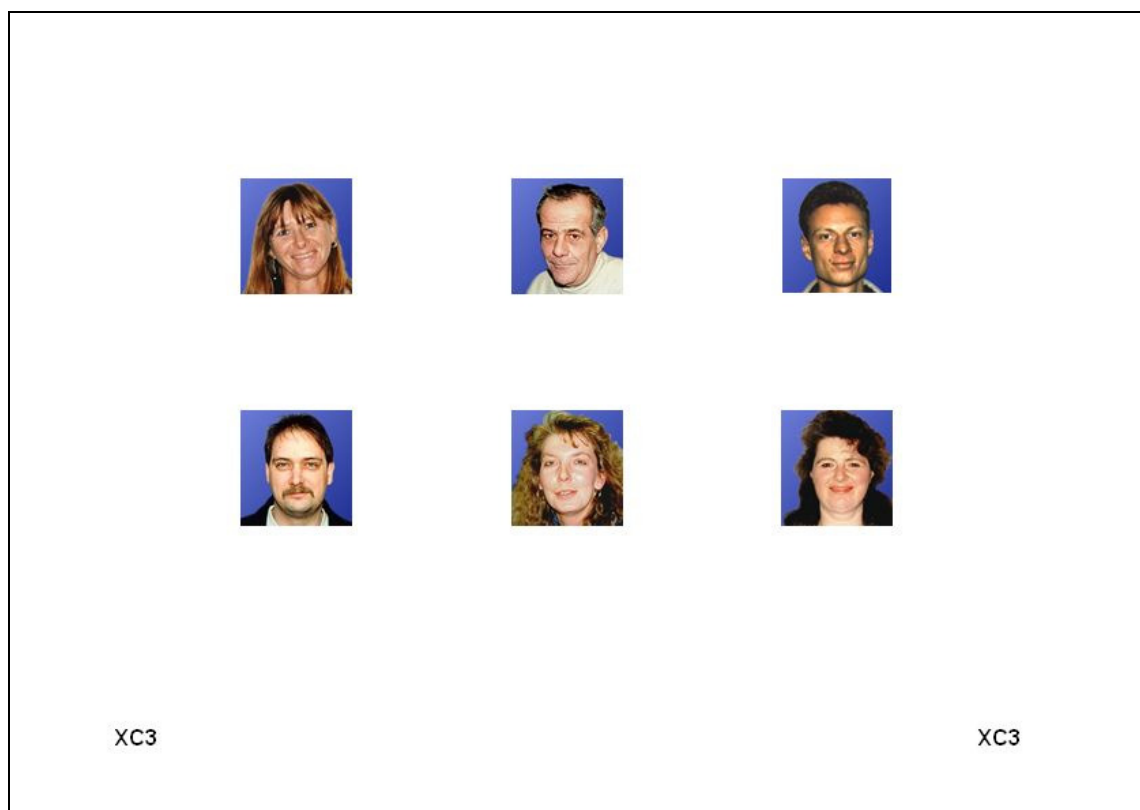


Bild 5: Vorderseite des Blattes XC3 des Gesichter-Gedächtnis-Tests zum Lernen von sechs Gesichtern, verkleinert vom farbigen Original aus A4-Format. Rückseite des Blattes siehe Bild 6.

Auf jedem Karton waren auf der Vorderseite sechs zu lernende Gesichter abgebildet (siehe Bild 5), jeweils drei Gesichter von Frauen und drei von Männern. Auf der Rückseite waren jeweils 24 Gesichter abgebildet (siehe Bild 6), unter die die sechs zu lernenden Gesichter gemischt waren. Dabei waren auf den mit dem gleichen Buchstaben bezeichneten Kartons die gleichen Gesichter, jedoch bei Messzeitpunkt 1, 2 und 3 in jeweils anderen Positionen zu sehen, siehe Seite 112ff. Die Bild-Dateien der Gesichter entstammten dem interaktiven Gedächtnis-Trainer der HJP-Multimedia GmbH, Furtwangen (<http://www.hjp-multimedia.de>), mit Genehmigung des Geschäftsführers Herrn Hans-Josef Pier.

Bei der Testkonstruktion waren die Gesichter auf beide Testversionen und die jeweils drei Karton-Versionen (A, B und C) zufällig verteilt worden. Ebenso waren die unterschiedlichen Gesichterarrangements für die jeweils drei Messzeitpunkte (1, 2 und 3) randomisiert erzeugt, wobei solange randomisiert wurde, bis auf der Rückseite nie mehr als je zwei der sechs zu lernenden Gesichter direkt nebeneinander lagen. So sollte eine etwa gleiche Schwierigkeit beider Testversionen gewährleistet werden. Durch die Randomisierung der Testversionsreihenfolge war gewährleistet, dass unterschiedliche Schwierigkeiten beider Versionen keinen systematischen Fehler erzeugen und nur die Störvarianz erhöhen würden. So konnte auf eine Schwierigkeitsvalidierung vor Studienbeginn verzichtet werden. Einen Schwierigkeitsvergleich der beiden Testformen X und Y des GGT enthält Tabelle 22 auf Seite 108 im Anhang dieser Arbeit.

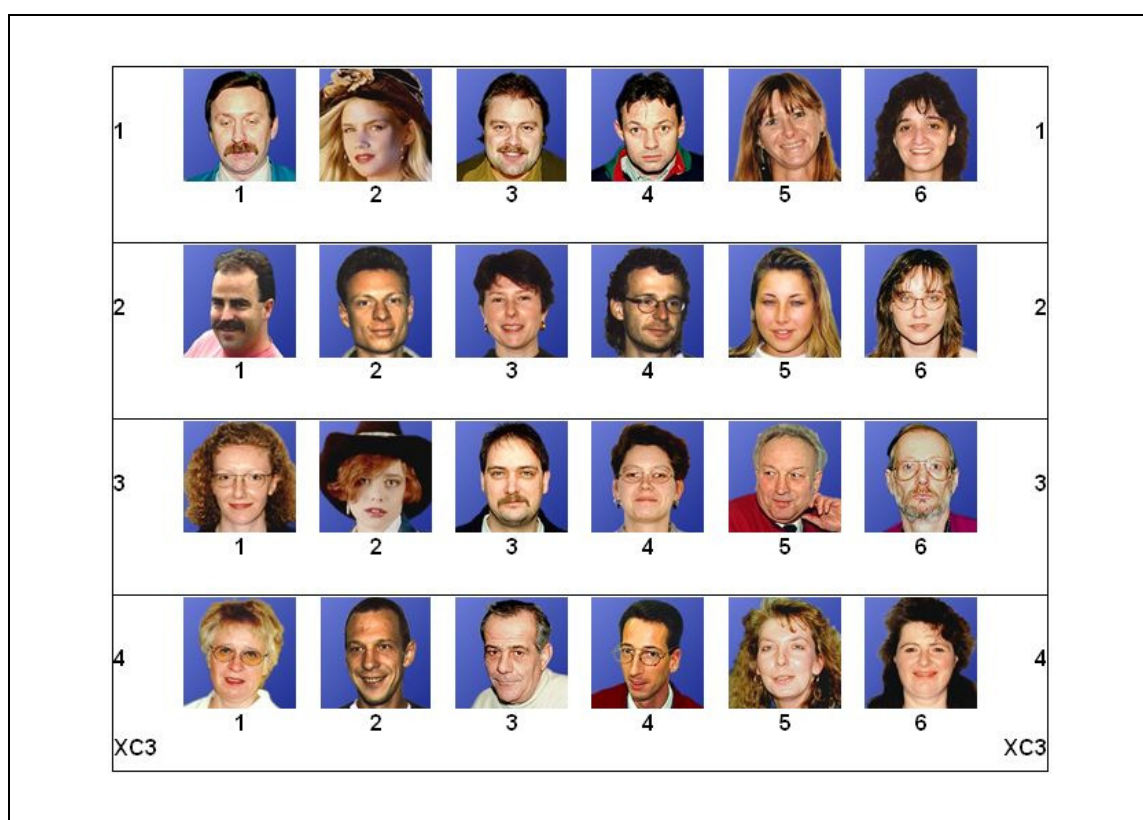


Bild 6: Rückseite des Blattes XC3 des Gesichter-Gedächtnis-Tests zum Wiedererkennen von sechs Gesichtern, verkleinert vom farbigen Original aus A4-Format. Vorderseite des Blattes siehe Bild 5.

Im ersten Messzeitpunkt einer Testsitzung wurde dem Patienten die Vorderseite von Karton A1 vorgelegt, auf der 6 zu lernende Gesichter abgebildet waren. Der Patient durfte diese beliebig lange betrachten (Bei unseren Patienten: Minimum etwa 10 Sekunden, Maximum etwa zwei Minuten, meistens ca 10 bis 30 Sekunden). Dann durfte der Patient den Karton umdrehen und musste die 6 Gesichter aus 24 Gesichtern wiedererkennen. Schließlich wurde der Patient aufgefordert, den Karton umzudrehen und das Ergebnis selbst zu überprüfen. Außerdem wurde er gebeten, sich die zu lernenden Gesichter noch einmal für die späteren Abfragen einzuprägen. Mit den Kartons 1B und 1C wurde ebenso verfahren. Im ersten Testdurchgang konnten somit 0 bis 18 Gesichter wiedererkannt werden.

Nach dem Nine-Hole-Peg-Test, etwa 20 Minuten nach dem Lerndurchgang, wurden dem Patienten nacheinander die Rückseiten der Kartons 2A, 2B und 2C vorgelegt und der Patient aufgefordert, die jeweils 6 gelernten Gesichter zu zeigen. Jetzt hatte der Patient keine Gelegenheit, sein Ergebnis zu überprüfen oder die Gesichter neu zu lernen. Ebenso im dritten Durchgang am Ende der Testsitzung mit den Kartons 3A, 3B und 3C, etwa 40 Minuten nach dem Lerndurchgang.

Der Gesamtscore des Gesichter-Gedächtnis-Tests wurde als Summenscore der 0 bis 54 richtig wiedererkannten Gesichter gebildet.

Mehrfachwahl-Wortschatz-Test (MWT)

Der Mehrfachwahl-Wortschatz-Test (MWT) ist ein Test zur Messung der allgemeinen kristallisierten Intelligenz anhand des im Laufe des Lebens angesammelten Wortschatzes. In 37 Wortreihen befindet sich jeweils ein echtes Wort neben vier ähnlichen erfundenen Wörtern. Das echte Wort ist jeweils anzukreuzen. Der Test beginnt mit einfachen Aufgaben und wird zunehmend schwieriger. Die richtig angekreuzten Wörter werden gezählt, so dass maximal 37 Punkte erreicht werden können.

- 32. Pucker – Keuper – Eucker – Reuspeck – Urkane
- 33. Spirine – Saprin – Parsin – Purin – Asprint
- 34. Kulon – Solgun – Koskan – Soran – Klonus
- 35. Adept – Padet – Edapt – Epatt – Taped
- 36. Gindelat – Tingerat – Indigenat – Nitgesaar – Ringelaar
- 37. Berkizia – Brekzie – Birakize – Brikazie – Bakiria

Bild 7: Ausschnitt aus dem Mehrfach-Wortschatz-Test (Version B).

Der MWT ist weitgehend unabhängig von der aktuellen Kurzspeicherkapazität der Probanden (Lehrl 1991). Der MWT lag in zwei Versionen (A und B) gleicher Schwierigkeit vor. Die Version vor der Operation wurde für jeden Patienten randomisiert. Im Follow-up wurde dann die jeweils andere Version verwendet. Einen Schwierigkeitsvergleich der beiden Testformen A und B des MWT enthält Tabelle 22 auf Seite 108 im Anhang dieser Arbeit.

Zahlengedächtnistest

Der Zahlengedächtnistest (Zahlenpyramide, siehe Bild 8 auf Seite 41, vollständig auf Seite 130) wurde dem Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE-R) entnommen (Tewes 1991). Es handelt sich um einen Test des Kurzzeitgedächtnisses für Zahlen. Der Testleiter liest eine Ziffernfolge vor (etwa eine Ziffer je Sekunde) die der Proband nachzusprechen hat. Nach je zwei Ziffernfolgen gleicher Länge verlängert sich die nächste Folge um eine Ziffer. Der Test wird abgebrochen sobald der Patient hintereinander zwei Folgen gleicher Länge nicht korrekt wiedergeben kann. In einer zweiten Testhälfte muss der Patient dann einen zweiten Set von Ziffernfolgen jeweils in umgekehrter Reihenfolge wiederholen. Für jede korrekt wiedergegebene Folge erhält der Proband einen Punkt. Vorwärts und rückwärts sind jeweils maximal 14 Punkte, insgesamt 28 Punkte erreichbar.

Zahlen vorwärts wiedergeben		2,1,0
1	5 - 8 - 2	
	6 - 9 - 4	
2	6 - 4 - 3 - 9	
	7 - 2 - 8 - 6	
3	4 - 2 - 7 - 3 - 1	
	7 - 5 - 8 - 3 - 6	
4	6 - 1 - 9 - 4 - 7 - 3	
	3 - 9 - 2 - 4 - 8 - 7	
5	5 - 9 - 1 - 7 - 4 - 2 - 8	
	4 - 1 - 7 - 9 - 3 - 8 - 6	
6	5 - 8 - 1 - 9 - 2 - 6 - 4 - 7	
	3 - 8 - 2 - 9 - 5 - 1 - 7 - 4	
7	2 - 7 - 5 - 8 - 6 - 2 - 5 - 8 - 4	
	7 - 1 - 3 - 9 - 4 - 2 - 5 - 6 - 8	
Summe:		max = 14

Bild 8: Zahlengedächtnistest: Ziffernfolgen vorwärts.

Standard Progressive Matrices (SPM) Intelligenztest

Der Standard Progressive Matrices Test ist ein Test der allgemeinen Intelligenz anhand des figürlich-logischen Denkens (K. A. Heller 1998). Der Proband soll geometrische Figuren oder Muster ergänzen, indem er aus jeweils 6 oder 8 Antwortalternativen die passende auswählt (siehe Beispielaufgabe Bild 9 auf Seite 42).

Um Übungeffekte zu minimieren und die Beantwortungsdauer zu verkürzen wurde der SPM-Test für diese Studie nach der im Testheft angegebenen Aufgabenschwierigkeit in zwei Testhälften gleicher Schwierigkeit geteilt. Jede Testhälfte bestand aus 31 Aufgaben ansteigender Schwierigkeit. In jeder der beiden Testversionen waren zwischen 0 bis 31 Punkte erreichbar. In für jeden Patienten randomisierter Reihenfolge wurde eine Testhälfte vor der Operation und die andere im Follow-up verwendet.

Die Aufteilung der Aufgaben des Original-SPM auf die beiden Testhälften ist in Tabelle 25 auf Seite 131 dokumentiert. Einen Schwierigkeitsvergleich der beiden Testformen A und B des SPM enthält Tabelle 22 auf Seite 108 im Anhang dieser Arbeit.

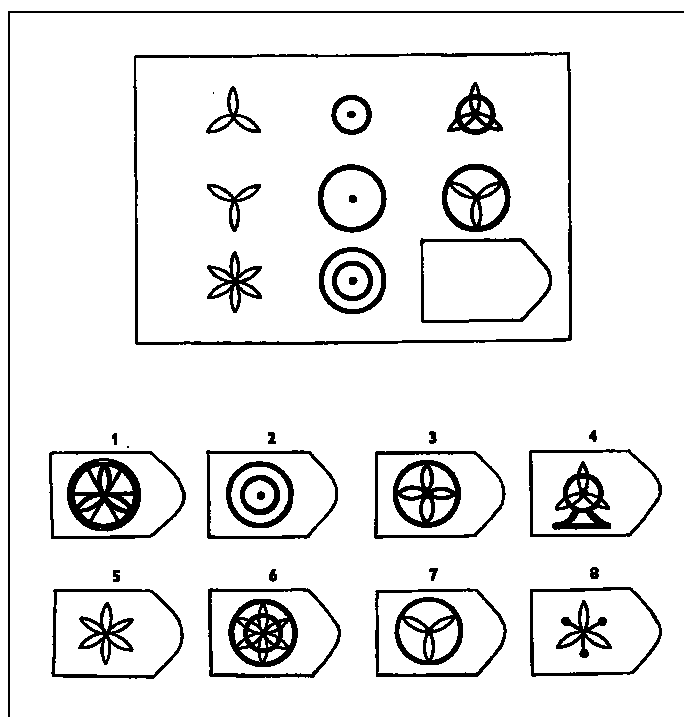


Bild 9: Eine Aufgabe aus dem SPM-Intelligenztest.

4.2.2 Neurologische Skalen

Folgende gut etablierte neurologische Skalen wurden vor der Operation und im Follow-up mindestens drei Monate nach der Operation erhoben:

National Institutes of Health Stroke Survey (NIHSS)

NIHSS: Die Skala des National Institutes of Health Stroke Survey (NIHSS) besteht aus 15 Subskalen zu Bewusstsein, Augenbewegung, Sehfeld, Gesichtsmotorik, Kontrolle der Extremitäten, Ataxie, Sensorik, Sprache, Dysarthrie und Neglect. Ein Gesamtscore von 0 steht für einen komplett unauffälligen Befund, das Maximum liegt bei 42. Bei einem einseitigen Schlaganfall mit Hemiparese, Hemiopie, Hemineglect und Aphasie liegt das Maximum bei 31 (T. Brott 1989; L. B. Goldstein 1989).

<p>1c. LOC Commands: The patient is asked to open and close the eyes and then to grip and release the non-paretic hand. Substitute another one step command if the hands cannot be used. Credit is given if an unequivocal attempt is made but not completed due to weakness. If the patient does not respond to command, the task should be demonstrated to him or her (pantomime), and the result scored (i.e., follows none, one or two commands). Patients with trauma, amputation, or other physical impediments should be given suitable one-step commands. Only the first attempt is scored.</p>	<p>0 = Performs both tasks correctly. 1 = Performs one task correctly. 2 = Performs neither task correctly.</p>
---	--

Bild 10: Ein Item aus dem National Institutes of Health Stroke Survey (NIHSS)

Rankin-Skala

Die Rankin-Skala ist eine sechsstufige Skala zur Beurteilung des Grades der Behinderung, von 0 (keinerlei Symptome) bis 5 (schwere Behinderung mit ständiger intensiver Pflegebedürftigkeit) (Rankin 1957; van Swieten 1988), siehe Bild 11.

Beschreibung	Score
keine Symptome	0
keine signifikante Behinderung, kann alltägliche Dinge selbst verrichten	1
leichte Behinderung, kann nicht an allen Aktivitäten des täglichen Lebens teilnehmen, sich aber weitgehend selbst helfen	2
moderate Behinderung, benötigt Hilfe im Alltag, kann ohne Hilfe gehen	3
stärkere Behinderung, kann nur mit Hilfe gehen, benötigt Hilfe in der Selbstversorgung	4
schwere Behinderung, bettlägerig, inkontinent, benötigt konstante Pflege und Aufmerksamkeit	5

Bild 11: Rankin-Skala zur Beurteilung des Behinderungsgrades.

Barthel-Index

Der Barthel Index bewertet 10 Aktivitäten des täglichen Lebens mit einem Gesamtscore von 100 (funktionelle Unabhängigkeit) bis 0 (völlige Abhängigkeit) (Collin 1988; Mahoney 1965), siehe Bild 12:

Item	Beschreibung/Score
Essen und Trinken	10=unabhängig 5=Hilfe beim Schneiden 0=abhängig
Transfer Bett/Stuhl	15=unabhängig 10=wenig Hilfe 5=alleine sitzen 0=abhängig
Körperpflege	5=unabhängig 0= abhängig
Toilettenbenutzung	10=unabhängig 5=wenig Hilfe 0=abhängig
Bad/ Duschen	5=unabhängig 0=abhängig
50m auf Ebene gehen	15=unabhängig 10=mit Hilfe 5=Rollstuhl 0=unmöglich
Treppensteigen	10=unabhängig 5=mit Hilfe 0=unmöglich
Anziehen (inkl. Schuhe)	10=unabhängig 5= mit Hilfe 0=abhängig
Darminkontinenz	10=kontinent 5=gelegentliche Inkontinenz 0=inkontinent
Harninkontinenz	10=kontinent 5=gelegentliche Inkontinenz 0=inkontinent

Bild 12:Barthel-Index zur Beurteilung der Unabhängigkeit bei Aktivitäten des täglichen Lebens..

Nine-Hole-Peg-Test (NHPT)

Die Geschicklichkeit der Hände wurde mit dem Nine-Hole-Peg-Test gemessen (Mathiowetz 1985). Der NHPT (siehe Bild 13) besteht aus einem quadratischen Holzbrett von 10cm Kantenlänge mit 9 in drei mal drei Reihen angeordneten Löchern, die jeweils 15mm auseinander liegen und 10mm Durchmesser haben und aus einer Schale in der 9 Holzdübel von 9mm Durchmesser und 32mm Länge liegen. Vor dem Start des Tests stand die Schale jeweils auf

der dem Patienten gegenüber liegenden Seite des Holzbrettes. Die Patienten mussten die Holzdübel je einmal mit der linken und einmal mit der rechten Hand möglichst schnell aus der Schale entnehmen und in die Löcher stecken. Die Zeit zwischen dem Wort „Start“ im Countdown („drei“, „zwei“, „eins“, „Start“) und dem Loslassen des letzten Dübels wurde in hundertstel Sekunden gestoppt. Die Patienten wurden instruiert, die Dübel nacheinander einzeln aus der Schale zu nehmen und in das Brett zu stecken. Das Entnehmen mehrerer Dübel wurde jedoch toleriert. In einigen wenigen Fällen wurde der Test bei groben Fehlern nicht gewertet und wiederholt, zum Beispiel als ein Dübel auf den Fußboden fiel. Als Gesamtscore für den NHPT wurde der Durchschnitt der Zeiten beider Hände verwendet.



Bild 13: Nine-Hole-Peg-Test.

4.2.3 Emotionaler Status

Der emotionale Status wurde vor der Operation und im Follow-up mindestens drei Monate nach der Operation mit folgenden Tests ermittelt:

Erlanger Depressivitäts-Skala (EDS)

Irvine (1998) fordert die Kontrolle der Depressivität bei Studien zur kognitiven Leistung von CEA-Patienten, um eine gesonderte Auswertung oder den Ausschluss depressiver Patienten zu ermöglichen.

	stimmt genau	fast richtig	nur halb richtig	stimmt kaum	ganz falsch
Ich möchte lustig sein					
Ich glaube, ich kann nichts leisten					
Ich möchte weinen					
Ich möchte etwas unternehmen					
Ich bin traurig					
Ich möchte mir etwas antun					
Ich fühle mich nicht wohl					
Ich mag mich selbst nicht mehr					
Ich kann mich nicht entscheiden					

Bild 14: Items der Erlanger Depressivitäts-Skala (EDS). Der Patient soll in jeder Zeile ein Antwort-Feld markieren.

Zur Messung der Depressivität wurde in unserer Studie die Erlanger Depressivitäts-Skala (EDS) verwendet (Lehrl 1983). Das erste Item ist ein Beispielitem. Es folgen acht Items, die ausgewertet werden. Die EDS ist damit kürzer als die meisten anderen Depressivitätstests. Es sind zwischen 0 und 32 Punkte erreichbar. Höhere Werte beschreiben eine höhere Depressivität. Tabelle 2 zeigt die Zuordnung von diagnostischen Klassifikationen zu den Punktzahlen aus dem EDS-Manual:

Punktzahl	depressive Symptomatik
0 - 5	keine
6 - 9	fraglich
10 - 16	leicht
17 - 24	mittel
25 - 32	schwer

Tabelle 2: Erlanger Depressivitäts-Skala (EDS): Punktzahl und Klassifikation (Lehrl 1983, S.19).

Visuelle Analogskalen (VAS) zur Selbstbeurteilung des emotionalen Status

Der Fragebogen mit visuellen Analogskalen (VAS, siehe Seite 110) zur Selbstbeurteilung des emotionalen Status wurde von mir selbst konstruiert. Der Patient beurteilte seinen Status auf 13 Analogskalen von 0 bis 80mm. Jede Skala wurde durch zwei gegensätzliche Adjektive an den Enden (zum Beispiel "krank" und "gesund") und eine dazwischen gedruckte 80mm lange Linie gebildet. Zur Vermeidung von Wertehäufungen und Deckeneffekten wurden die Patienten instruiert, nicht die Extremwerte anzukreuzen, sondern zwischen den beiden Enden der Skala. Der Gesamtscore wurde als Mittelwert gebildet und nimmt damit einen Wert von 0 bis 80 an. So wurde die Vergleichbarkeit von einzelnen Skalen mit dem Gesamtscore erleichtert.

traurig	_____	fröhlich
abhängig	_____	selbständig
untätig	_____	aktiv
sorgenvoll	_____	zuversichtlich
schläfrig	_____	wach

Bild 15: Ausschnitt aus den Visuellen Analogskalen (VAS) zur Selbstbeurteilung von Stimmung, Leistungsfähigkeit und Gesundheit. Der Vollständige Test steht im Anhang auf Seite 110. Im Original sind die einzelnen Skalen genau 80mm lang.

Bei den selbst konstruierten visuellen Analogskalen wurde in einer späteren Auswertung die Zulässigkeit der additiven Zusammenfassung durch positive Korrelationen der Skalen mit dem Gesamtscore belegt, siehe Tabelle 24, Seite 111.

4.2.4 Begleiterkrankungen

Informationen über die Begleiterkrankungen der Patienten lieferte der behandelnde Oberarzt der Neurologie, Herr PD Dr. Hartmann. Zusätzlich wurden die Patienten befragt. Bei unvollständigen oder widersprüchlichen Informationen wurden die Arztbriefe und Patientenakten konsultiert.

- Periphere arterielle Verschlusskrankheit (pAVK): 0 = nein, 1 = pAVK nach Aktenlage und Befragung des Patienten bei Claudicatio oder bei peripheren Revaskularisationen oder Amputationen. In jedem Fall wurden die peripheren Pulse getastet. Im Zweifel oder bei neuem Verdacht auf pAVK wurde der Knöchel-Arm-Index (ABI) für beide Seiten bestimmt als Quotient aus systolischem Blutdruck am Fußgelenk und systolischem Blutdruck am Handgelenk. Unterschritt der ABI-Index den Grenzwert von 0,9, so wurde dies als pAVK gewertet.
- Koronare Herzkrankheit: 0 = nein, 1 = koronare Herzkrankheit in der Anamnese.
- früherer Herzinfarkt: 0 = nein, 1 = früherer Herzinfarkt in der Anamnese.
- Vorhofflimmern: 0 = nein, 1 = Vorhofflimmern aus der Anamnese oder aus dem EKG, das bei jedem Patienten durchgeführt wurde.

4.2.5 Risikofaktoren

Informationen über die potentiellen Risikofaktoren der Patienten lieferte der behandelnde Oberarzt der Neurologie, Herr PD Dr. Hartmann. Zusätzlich wurden die Patienten befragt. Bei unvollständigen oder widersprüchlichen Informationen wurden die Arztbriefe und Patientenakten konsultiert.

- Rauchstatus: 2 = Raucher, 1 = Exraucher seit mindestens einem Jahr, 0 = Nichtraucher = Nieraucher; im Follow-up: 2 = Raucher, 0 = Nichtraucher oder Exraucher seit mindestens einem Monat.
- Rauchjahre: Die Anzahl der Lebensjahre in denen geraucht wurde.
- Rauchaufgabe: Die Anzahl der Jahre seitdem das Rauchen aufgegeben wurde.
- arterielle Hypertonie: 0 = nein, 1 = Blutdruck ab 140/90 mm Hg oder Anwendung von blutdrucksenkender Medikation (Deutsche Hochdruckliga 2003).
- Diabetes mellitus: 0 = nein, 1 = chronische Hyperglykämie, die eine Diät, orale Medikation oder Insulingaben notwendig machte oder mehrfach erhöhte Nüchtern-Blutzuckerspiegel.
- Hypercholesterinämie: 0 = nein, 1 = Serumcholesterinspiegel von über 200mg/dl oder Medikation

4.2.6 Stenosegrade und hämodynamische Indikatoren

- Stenosegrade in Prozent (links und rechts, vor OP und im Follow-up). Die Stenosegrade wurden wie bei Arning (2003) beschrieben in einer Kombination aus direkten und indirekten hämodynamischen Ultraschall-Kriterien erhoben.
- Pulsatilitätsminderung = abgeflachte Pulswellenform = reduzierte systolischer Flussgeschwindigkeit und Pulsatilität in der Arteria Cerebri Media im transkraniellen Duplex oder Doppler (TCD) (Hartmann 2000):
0 = nein, 1 = ja (jeweils links oder rechts, vor OP und im Follow-up).
- Crossflow = im transkraniellen Duplex oder Doppler festgestellter Kollateralfluss von einer Seite des Circulus Willisii zur anderen über die Arteria Communicans Anterior oder Posterior:
0 = nein, 1 = ja (jeweils nach links oder nach rechts, vor OP und im Follow-up).

Bei allen Patienten wurde vor der Operation zur Bestätigung der Stenose und für die Detektion möglicher anderer Stenosen stets eine extra- und intrakranielle MRT-Arteriographie mit dem Siemens Vision MRT durchgeführt.

Zur besseren Reliabilität der Vor- und Nach-OP Stenosegrade wurden ausschließlich die in der Ultraschall-Untersuchung vom behandelnden Oberarzt der Neurologie, Herrn PD Dr. Hartmann, bestimmten Stenosegrade verwendet. Die Stenosegrade wurden bestimmt mit dem DWL Multidop X4 Dopplergerät mit 4 MHz und 8 MHz Stiftsonden. In den meisten Fällen wurden die Stenosegrade mit dem ATL HDI5000 Duplexgerät mit 5-12MHz Linearschallkopf überprüft.

4.2.7 Ischämien als Folgen von Carotisstenosen

Die Diagnosen der Ischämien stammten aus der Anamnese oder aus Untersuchungen in oder auf Veranlassung unserer neurologischen Abteilung.

- Schlaganfall: 0 = nein, 1 = ja
- TIA: 0 = nein, 1 = ja

- permanente ischämische Sehstörung: 0 = nein, 1 = ja
- Amaurosis fugax: 0 = nein, 1 = ja

4.2.8 Operation der Carotis-Arterie

- Seite der Operation: links oder rechts
- Operationsmethode
- Anwendung eines intraoperativen Shunts
- Abklemmzeit der Arterie
- Änderungen der somatosensorischen evozierten Potentiale (SEP) im EEG während der CEA: 0 = Änderung < 50%, 1 = Änderung \geq 50%

Die Details zur Operationsdurchführung wurden aus den Operationsprotokollen erhoben.

4.3 **Datenerhebung**

4.3.1 Räume und Termine

Mit dem Raum HH4211 im Hauptgebäude des Klinikums stand ein Büro zur Verfügung, in dem die Tests weitgehend störungsfrei durchgeführt werden konnten. Zur Vereinheitlichung der Versuchsbedingungen wurden fast alle Tests in diesem Raum durchgeführt. Die Ultraschalluntersuchungen wurden im unmittelbar daneben liegenden Doppler-Labor durchgeführt.

Vor CEA erfolgten die Testsitzungen meistens nach Aufnahme des Patienten zur Operation. In der Regel besuchte ich die Patienten in Ihrem Krankenzimmer und vereinbarte einen Termin für den gleichen oder nächsten Tag, zu dem sie zur Testsitzung erschienen. Häufig waren die Patienten auch bereit, die Tests sofort durchzuführen. Die Ultraschalluntersuchung war in der Regel schon durchgeführt.

Zur Durchführung des Follow-up wurden alle Patienten telefonisch kontaktiert. Dies erfolgte in Absprache mit dem behandelnden Neurologen, Herrn PD Dr. Hartmann, der die neurologischen Untersuchungen und die Ultraschalldiagnostik durchführte. Einige Patienten meldeten sich auch selbst, um einen Termin zu vereinbaren. Im Follow-up wurde die Ultraschalluntersuchung meistens unmittelbar vor oder nach der Testsitzung durchgeführt.

4.3.2 Erfassung der Rohdaten

Um von technischen und räumlichen Gegebenheiten möglichst unabhängig zu sein, wurden ausschließlich einfach durchführbare Papier-Bleistift-Tests ausgewählt, so dass notfalls auch auf andere Räume ausgewichen werden konnte.

Bei der Testdurchführung wurden alle Antworten und Ergebnisse auf Papier festgehalten, teilweise vom Testleiter (Risikofaktoren, Mini-Mental, Gesichter-Gedächtnis-Test, Zahlenp-

ramide, Nine-Hole-Peg-Test), teilweise von den Probanden selbst (Visuelle Analogskalen, Erlanger Depressions-Skala, Mehrfachwahl-Wortschatz-Test, SPM Intelligenztest)

Die Rohdaten wurden noch während der jeweiligen Testsitzung oder unmittelbar darauf zu Gesamtscores für die einzelnen Tests zusammengefasst und in eine Excel-Tabelle übernommen.

4.3.3 Testablauf

Die Testsitzung wurde in folgender Reihenfolge gestaltet:

1. Zunächst wurde der Patient informiert über Zweck der Studie, Charakter der Tests, Follow-up-Absicht, Freiwilligkeit der Teilnahme, Art der Datensammlung und Datenanonymisierung.
2. Der Patient erhielt dann eine schriftliche Information (siehe Seite 106) mit Erklärung zur Freiwilligkeit der Teilnahme, Genehmigung der Datenerhebung und Datenverwendung. Die teilnehmenden Patienten unterschrieben diese Einverständniserklärung (siehe Seite 107) und wurden jeweils aufgefordert Adresse und Telefonnummer zwecks Follow-up-Kontaktaufnahme zu ergänzen.
3. Eine Kopie der Patienteninformation und der Einverständniserklärung wurde dem Patienten ausgehändigt. Das unterschriebene Original der Einverständniserklärung mit Patientennummer und Adresse des Patienten wurde von den Antwortbögen getrennt aufbewahrt.
4. Der Patient wurde befragt zu Schulbildung, Händigkeit, Risikofaktoren, Begleiterkrankungen, Stenosegraden, Operationsseite (links oder rechts) und geplantem Operationstermin. Das Gespräch wurde im Sinne der Sekundärprävention von Schlaganfällen dazu genutzt, die Patienten auf die kontrollierbaren Risikofaktoren Hypertonie, Diabetes mellitus und Rauchen aufmerksam zu machen. Ebenso wurden Schlaganfallsymptome und die Notwendigkeit einer unverzüglichen Reaktion angesprochen. In der Regel zeigten sich die Patienten über diese wichtigen Verhaltensregeln schon aus Arztgesprächen aufgeklärt.
5. Mini-Mental-Test
6. Selbstbeurteilung (13 visuelle Analogskalen, VAS)
7. Erlanger Depressions-Skala
8. Gesichter-Gedächtnis-Test: erster Teil: drei mal je sechs Gesichter lernen und jeweils unmittelbar darauf aus je 24 Gesichtern wiedererkennen (maximal 18 Punkte, randomisiert Form A oder B)
9. Mehrfachwahl-Wortschatz-Test (randomisiert Form A oder B)
10. Zahlengedächtnis-Test
11. Nine-Hole-Peg-Test

12. Gesichter-Gedächtnis-Test: zweiter Teil: Gesichter wiedererkennen, drei mal sechs Gesichter (maximal 18 Punkte)
13. SPM Intelligenztest (randomisiert Form A oder B)
14. Gesichter-Gedächtnis-Test: dritter Teil: Gesichter wiedererkennen, drei mal sechs Gesichter (maximal 18 Punkte)

4.3.4 Follow-up: Zeitpunkt und Durchführung

Verschlechterte kognitive Leistungen sind in den ersten vier Wochen nach CEA berichtet worden, siehe Tabelle 1 auf Seite 25. Für größere Zeitabstände wurden keine Verschlechterungen der mittleren Leistungen von Patientengruppen berichtet. Der am häufigsten verwendete Follow-up-Abstand von drei Monaten dürfte deshalb ausreichen, um vorübergehende negative Effekte der Operation oder Narkose zu umgehen. Deshalb wurde für diese Studie ein Mindestabstand des Follow-up von drei Monaten nach der Operation angestrebt. Dieser Abstand gilt auch als ausreichend, um Übungseffekte bei Testwiederholung zu minimieren (Irvine 1998, p.197). Wie viel Zeit cerebrale Erholungseffekte infolge verbesserter Perfusion benötigen, ist unklar. Nach Erfahrungen mit Schlaganfallpatienten und Hirnverletzten sind Verbesserungen in den ersten Monaten am häufigsten, aber auch später noch möglich. Im fortgeschrittenen Alter muss jedoch auch mit der Entwicklung von Demenzen gerechnet werden. Deshalb und zur Gewährung einer ausreichend einheitlichen Datenbasis wurden die Follow-up-Messungen möglichst innerhalb eines Jahres nach der Operation durchgeführt.

Am Beginn der Follow-up-Sitzung wurden die Patienten anhand eines Follow-up-Fragebogens mündlich zu den Operationsfolgen ("Wie lange hat es gedauert, bis Sie sich von der Operation erholt haben?" und "Gab es Probleme mit der Wundheilung?") und ihrer Operations-Zufriedenheit befragt ("Würden Sie sich noch einmal operieren lassen?"). Außerdem wurde anhand einer Symptomliste geklärt, ob seit der Operation eine Ischämie (Schlaganfall, TIA oder Amaurosis) aufgetreten war. Das gegenwärtige Rauchverhalten wurde erhoben. Raucher und Exraucher wurden auf die Bedeutung des Rauchens als Risikofaktor hingewiesen und alle Patienten wurden auf die Notwendigkeit einer unverzüglichen Reaktion im Falle von Ischämie-Symptomen aufmerksam gemacht.

Anschließend wurden die Tests beginnend mit dem Mini-Mental-Test in der gleichen Reihenfolge wie vor der Operation durchgeführt. Bei den Tests, die in zwei Versionen vorlagen, wurde im Follow-up für jeden Patienten die jeweils andere Version verwendet.

Die neurologische Untersuchung inklusive Ultraschalldiagnostik wurde im Follow-up meistens unmittelbar vor oder nach den psychologischen Tests vom behandelnden Oberarzt der Neurologie, Herrn PD Dr. Hartmann, durchgeführt.

4.4 Statistik

4.4.1 Stichprobenumfang

Bei den von Lunn (1999) gesammelten 28 Studien zu kognitiven Veränderungen nach Carotisendarterektomie lag die Stichprobengröße im Median bei 30. Der Median der Stichprobengröße der Studien, die Verbesserungen berichteten, lag bei 28 (siehe Seite 24).

Deshalb wurde geplant, in unsere Studie 40 zur Carotisendarterektomie aufeinander folgende Patienten einzuschließen. Dies sollte eine Reserve für den Ausfall von Patienten zum Follow-up lassen und dennoch nach den Ergebnisse der anderen Studien ausreichen, um Veränderungen nach CEA nachweisen zu können. Durch im Vergleich zu vielen anderen Studien sorgfältigere Testauswahl, hohe Skalenniveaus und kontrollierte Versuchsbedingungen sollten maximale Teststärken und minimale Störvarianzen erreicht werden.

Die Stichprobengröße ist bei einer geringen Probandenzahl und der Betrachtung einer Vielzahl von Variablen stets ein Problem. Dem wurde durch mehrere Maßnahmen begegnet:

- Konzentration auf wenige Hauptfragestellungen mit Hypothesen.
- Die Ergebnisse aller kognitiven Leistungstests wurden zu einem Mittelwert der Prozentrangwerte der Einzeltests zusammengefasst. Dieser Gesamtscore der kognitiven Leistung wurde dann für die Auswertungen der Hauptfragestellung herangezogen. So konnten die Störvarianzen verringert und Probleme der Interpretation bei einer großen Zahl von statistischen Tests vermieden werden.
- Nebenfragen und Einzeltests sollten nur zusätzlich explorativ ausgewertet und entsprechend vorsichtig interpretiert werden, weil Zufallsbefunde hier aufgrund der höheren Zahl von Berechnungen wahrscheinlicher vorkommen.
- Auf multivariate Auswertungen wurde wegen fehlender Notwendigkeit zur Klärung der zentralen Fragen und angesichts der dafür kleinen Stichprobe verzichtet.
- Die Ergebnisse sollten im Rahmen der Ergebnisse anderer Studien diskutiert werden.

4.4.2 Zusammenfassungen von Daten

Aus zwei Gründen wurden Zusammenfassungen von Ergebnissen mehrerer Tests oder Untersuchungen vorgenommen: Erstens sollte die Störvarianz durch diese Zusammenfassung von mehr Items gesenkt werden. Zweitens sollten Zufallstreffer vermieden werden, die bei Durchführung einer größeren Zahl von Signifikanztests zu erwarten wären. Solche Zufallsergebnisse wären nur durch Korrekturmaßnahmen wie Bonferroni-Korrektur kontrollierbar, die die Teststärke wesentlich beeinträchtigten.

Um die Zusammenfassung zu ermöglichen, wurden zunächst für jeden Wert der Variablen ein Prozentrangwert im Vergleich zu allen anderen Werten vor und nach der Operation gebildet. Die Richtung einiger Skalen wurde dabei umgekehrt, so dass ein höherer Prozentrang immer ein besseres Ergebnis bedeutet. Zusammenfassungen wurden dann als Mittelwert der zusammenfassenden Variablen je Person und Messzeitpunkt gebildet.

Zur Auswertung der Hauptfragestellungen wurden so folgende Variablen zusammengefasst:

Zu einer Variablen "Kognitive Leistung" wurden die Ergebnisse aller fünf kognitiven Leistungstests zusammengefasst: Mini-Mental, Mehrfachwahl-Wortschatz-Test, Zahlengedächtnistest, Gesichter-Gedächtnis-Test, Standard Progressive Matrices.

Zu einer Variablen "Emotionaler Status" wurden die Ergebnisse der Erlanger Depressions-Skala und der visuellen Analogskalen zusammengefasst. Der Gesamtscore der visuellen Analogskalen wurde zuvor als Mittelwert aller Analogskalen gebildet. Der Prozentrangwert der Erlanger Depressions-Skala wurde durch absteigende Prozentränge, gebildet: Während bei der EDS ein höherer Wert höhere Depressivität bedeutet, bedeutete dann beim Emotionalen Status ein höherer Wert eine geringere Depressivität und bessere Stimmung und Selbsteinschätzung.

Zu einer Variablen "Neurologischer Status" wurden die Variablen Rankin, NIH Stroke Scale, Barthel Index und Nine-Hole-Peg-Test so zusammengefasst, so dass ein höherer Wert in der zusammengefassten Variablen ein besseres Ergebnis bedeutet.

Zu einer Variablen "Hämodynamische Auffälligkeit" wurden die Variablen Crossflow und Pulsatilitätsminderung unabhängig von der Seite zusammengefasst. Dies weil Patienten mit komplett verbesserter Hämodynamik mit den anderen Patienten bezüglich der Veränderung Ihrer kognitiven Leistung nach CEA verglichen werden sollten. Eine weitere Differenzierung hätte bei der relativ kleinen Patientenzahl die Sensibilität der Auswertung deutlich verringern können.

4.4.3 Starke Veränderungen bei einzelnen Personen

Maße der zentralen Tendenz, zum Beispiel Mittelwert oder Median, bergen die Gefahr von Fehlinterpretationen, weil unter Umständen selbst extrem unterschiedlich verteilte Daten zum gleichen Wert führen können.

Dennoch sind Maße der zentralen Tendenz sinnvoll, weil sie bei hohem Skalenniveau ermöglichen, auch geringe Gruppenunterschiede zu finden.

Es besteht aber die Gefahr der Zusammenfassung von gegensätzlichen Entwicklungen: "*A group mean score will reflect both the individuals whose performance is enhanced as a consequence of practice and those whose performance is compromised due to postoperative neurobehavioral dysfunction.*" (Murkin 1997, p.904). Gleiche Änderungswerte können außerdem entstehen durch eine Zusammenfassung von praktisch unbedeutenden kleinen Veränderungen bei vielen Patienten oder durch sehr große Veränderungen bei wenigen Patienten.

Einen ersten Schutz gegen Fehlinterpretationen bietet schon die genaue Sichtung der Verteilungen und die Anwendung von möglichst nonparametrischen statistischen Verfahren.

Collie (2002) propagiert den Einsatz von Maßen für starke individuelle Veränderungen in Studien zu Veränderungen der kognitiven Leistungsfähigkeit nach Herzoperationen. Indem

man zunächst eine Schwelle für starke Veränderungen definiert, werde man individuell bedeutsamen Veränderungen gerechter.

In Anlehnung an ein von Collie propagiertes *Kriterium für eine signifikante Veränderung* wurde für unsere Studie folgende Definition vorgenommen: Die Veränderung eines Patienten im Vergleich von Vor- und Nach-CEA-Test ist stark, wenn der Betrag der Veränderung eine Standardabweichung der Messungen vor der Operation überschreitet. Sinnvoll ist es, die Veränderung des Patienten auf einer dreiwertigen Skala abzubilden: +1 für eine starke Verbesserung, 0 für nicht stark verändert und -1 für eine starke Verschlechterung.

Als *Kriterium für eine starke Abweichung* eines Testwertes eines einzelnen Patienten von der durch den Mittelwert aller Patienten gebildeten Norm vor der Operation wurde die Abweichung um mindestens zwei Standardabweichungen definiert. Im Gegensatz zum Veränderungsmaß, wo der Patient an seiner eigenen Norm gemessen wird, sollte der Vergleich mit dem Gruppenmittelwert eine größere Varianz zulassen. Abweichungen von mindestens zwei Standardabweichungen haben bei Normalverteilung der Daten eine Wahrscheinlichkeit von knapp 5%. Sinnvoll ist es, das Ergebnis auf einer dreiwertigen Skala abzubilden: +1 für ein stark besseres Ergebnis, 0 für nicht stark abweichend und -1 für ein stark schlechteres Ergebnis.

Durch Anwendung dieser Datentransformationen wird das jeweilige Skalenniveau erheblich reduziert und kleine Veränderungen oder Abweichungen werden nicht mehr berücksichtigt. Deshalb wird die Auswertung der so transformierten Daten nur zusätzlich zu einer herkömmlichen Auswertung durchgeführt. Falls sich nach der Datentransformation ausreichende Zahlen von stark veränderten Patienten ergeben, sollten Zusammenhänge mit anderen Variablen exploriert werden.

4.4.4 Statistische Tests

Inferenzstatistiken wurden mit Excel 97, SPSS 11 oder SISA (Januar 2004) berechnet. SISA (Simple Interactive Statistical Analysis) ist ein webbasiertes Statistiktool von Daan G. Uitenbroek unter der Internet-Adresse <http://home.clara.net/sisa/>.

Die Korrektheit der Berechnungsverfahren wurde jeweils mindestens anhand eines Beispiels aus der Literatur oder eines Vergleichs von Ergebnissen aus einer anderen Software (Excel, SPSS oder SISA) validiert.

Es wurde eine SPSS-Datei zur Steuerung des Datenimports in SPSS und zur Ergänzung mit Variablen- und Wertebezeichnungen angelegt. Alle Variablen-Zusammenfassungen und auf Zusammenfassungen beruhenden Statistiken wurden mit SPSS berechnet.

Mit SPSS wurden Häufigkeiten, Mittelwerte, Standardabweichungen, t-Tests Levene-Tests und Korrelationen berechnet. Bei deutlichen Abweichungen von der Normalverteilung oder kleinem Stichprobenumfang wurden statt dem t-Test der Wilcoxon- oder der Mann-Whitney-Test mit SPSS berechnet. Als zweiseitige Wahrscheinlichkeiten wurden beim Mann-Whitney-

Test die doppelten einseitigen Wahrscheinlichkeiten angegeben. Bei den nonparametrischen Berechnungen wurde immer die Exact-Methode von SPSS genutzt.

Bei deutlichen Abweichungen von der Normalverteilung oder kleinem Stichprobenumfang wurde statt der Pearson-Korrelation die Spearman-Korrelation verwendet, die auf Rangdaten beruht. Der zum Vergleich von Varianzen verwendete Levene-Test gilt als robust gegen Abweichungen von der Normalverteilung.

Mit Excel wurden Häufigkeiten und der Binomial-Test berechnet.

Odds-Ratio und Fisher-Test wurden mit SISA oder mit SPSS berechnet.

Für den Fall, dass eine Hypothese nicht bestätigt werden könnte, der gefundene Effekt aber in die postulierte Richtung geht, sollte auf Grundlage der gefundenen Daten eine Poweranalyse durchgeführt werden. Die Poweranalyse sollte festzustellen, wie viele Patienten zu untersuchen wären, um den Effekt mit 80%iger Wahrscheinlichkeit (=Power) und 5%iger Irrtumswahrscheinlichkeit bestätigen zu können. Power-Berechnungen sollten mit dem SISA- und zur Kontrolle mit dem UCLA-Power-Calculator durchgeführt werden. Der UCLA-Power-Calculator ist ein webbasiertes Statistiktool der University of California unter der Internet-Adresse <http://calculators.stat.ucla.edu/>.

Für einseitige Hypothesen wurden einseitige Wahrscheinlichkeiten berechnet, ansonsten zweiseitige. Bei den im Ergebnisteil genannten Wahrscheinlichkeiten handelt es sich um zweiseitige Wahrscheinlichkeiten, sofern es nicht anders angegeben ist.

Hypothesentests

Die Hypothese verbesserter kognitiver Leistungen nach CEA sollte mit dem nonparametrischen Wilcoxon-Test einseitig geprüft werden. Der Wilcoxon-Test berücksichtigt dabei die Rangreihe der Beträge der positiven und negativen Differenzen der kognitiven Leistungen der einzelnen Patienten.

Die Hypothese verbesserter kognitiver Leistungen für hämodynamisch verbesserte Patienten sollte durch den Vergleich der Differenzen der kognitiven Leistungen (Nach-CEA minus Vor-CEA) der hämodynamisch verbesserten mit den übrigen Patienten per Mann-Whitney-Test einseitig geprüft werden. Der Mann-Whitney-Test ist ein nonparametrischer Test und vergleicht die Rangplätze der Werte beider Gruppen.